EVALUATION OF TOBACCO GRADING SYSTEMS BY MULTIVARIATE ANALYSIS OF THEIR CHEMICAL QUALITY PARAMETERS

EVALUATION DES SYSTEMES DE CLASSIFICATION DE TABACS PAR L'ANALYSE MULTIVARIEE DE LEURS PARAMETRES DE QUALITE CHIMIQUE

L.R. Vargas, J.I. Castaño and F.J. Palacio
Cía. Colombiana de Tabaco S.A., Investigación y Desarrollo
A.A. 828, Medellín, Colombia

SUMMARY

Three multivariate techniques were considered to evaluate different types and grades of tobacco using the results of nine chemical quality parameters.

Samples of tobacco already graded according to conventional criteria were submitted to a discriminant analysis to evaluate the operativity of the grading systems. Through Mahalanobis distances an order of similarity between types and grades was deduced, thus enabling the search of compatible tobaccos with similar chemical properties. By principal component analysis the system of nine variables was reduced to three components, in such a way, the result of the discriminant analysis was reproduced at 94%. A data file from different kinds of tobacco was analyzed through a non-hierarchical cluster method, which proved the great capacity of this technique to gather objects of similar chemical nature in homogeneous groups, and its advantage in the development of better grading criteria.

The application of these techniques involving physical and organoleptical parameters in addition to the chemical ones, will provide a powerful tool for future studies in this field.

RESUME

On a employé trois techniques multivariantes pour évaluer les différents types et degrés de tabacs en utilisant les résultats de neuf paramètres chimiques de qualité.

En partant d'échantillons classifiés au préalable selon des critères physiques conventionnels, on a fait usage d'une analyse discriminante pour évaluer l'opérativité des systèmes de classification. En utilisant les distances de Mahalanobis, on a pu déduire l'ordre de similarité entre les divers types et degrés permettant la recherche de tabacs compatibles selon leurs propriétés chimiques. En partant des composantes principales, le système de neuf variables a pu être réduit a trois variables, ainsi, les résultats de l'analyse discriminatoire ont été reproduits à 94%. Une méthode cluster non hiérarchique a été appliquée à plusieurs échantillons de différentes classes de tabacs ayant des propriétés chimiques similaires. Cette technique s'est montrée d'un grand potentiel pour le développement de meilleurs critères de classification.

L'application de ces méthodes, qui incluent des paramètres organoleptiques et physiques en plus des paramètres chimiques, pourrait devenir un outil puissant pour des recherches postérieures.

INTRODUCTION

The multivariate analysis is a statistical technique for data analysis that allows to work simultaneously with several variables, giving the possibility of an integral analysis of the information. Among the statistical techniques of this type, we can mention the Discriminant Analysis, the Principal Component Analysis (PC) and the Cluster Analysis.

The discriminant analysis is a supervised technique that starts from classified observations to determine whether or not an object belongs to an assigned population. The objective of the principal component analysis is to reduce the original variables to a few components without a significant loss of information.

The cluster analysis is a non-supervised classification technique which allows to group the objects according to some similarity criterion.

In this study, the foregoing techniques were applied for analysis of the tobacco grading system, the comparison between crops and the definition of equivalences between different grades of tobacco. A non-hierarchical cluster analysis was carried out with samples of different varieties so as to gather sample groups having similar chemical characteristics.

EXPERIMENTAL

a. Sample Identification

For the evaluation of the grading system, seven grades of the 1988 Flue-cured tobacco crop were classified according to conventional criteria (color, body, size, stalk position, etc.). The comparison of crops was made between 1986, 1987 and 1988 Flue-cured tobacco crops (Table 1). To group tobacco samples according to their chemical similarities, samples of Flue-cured, Burley, Black air-cured and another type of air-cured called Extra, were employed (Table 2).

Table 1. Sample identification. Seven Flue Cured Grades (A-G). 1986, 1987, 1988 Crops

1986	n	1987	n	1988	n	Description
A6	5	A7	9	A8	9	Cutters
B6	5	B7	8	B8	8	Leaf
C6	8	C7	9	C8	9	Lugs
D6	6	D7	9	D8	9	Smoking Leaf
E6	6	E7	9	E8	9	Primings
F6	7	F7	9	F8	9	Upper Leaf
G6	5	G7	9	G8	9	Mixed-non descript

n: number of samples analyzed

Table 2. Sample Identification (ID) for Non-Hierarchical Cluster Analysis

ID	n	Description
H	10	Leaf, Extra air-cured
I	10	Leaf, Black air-cured
J	10	Leaf, Flue-cured
K	9	Stem, Flue-cured
L	10	Leaf, Burley

n: number of samples analyzed

b. Chemical Quality Parameters

The parameters used as quality criteria were: Reducing Sugars, Total Nitrogen, Nicotine, Total Volatile Bases (TVB), Petroleum Ether Extracts (PEE), Ashes,

Chlorides and Sugar-Nicotine (Sug-Nic), Nitrogen-Nicotine (Nit-Nic) ratios. The chemical analyses were carried out according to CORESTA, ISO and other standard methods used for tobacco.

c. Data Analysis

The following computer software were utilized for the statistical analysis of the information: ·Systat 4.1, Statgraphics 4.0, Minitab 6.1, Lotus 2.2, Asym, and Harvard Graphics 2.13.

RESULTS AND DISCUSSION

a. Evaluation of Tobacco Grading Systems

This evaluation was made using discriminant analysis applied to seven grades of the 1988 Flue-cured tobacco crop. A preliminary variance analysis showed that the nine chemical parameters considered had a discriminant capability. The criteria for the classification were Mahalanobis' minimum distances (D^2_{iz}), according to the following definition:

Classify Z into πi, if:
$D^2_{iz} = \min \{D^2_{1z}, D^2_{2z}, \ldots, D^2_{kz}\}$; where i = 1,.,.,., 7, z = 1, .,.,., 62

The results are shown in Table 3, where the "total" row represent the number of pre-classified samples in each group according to conventional criteria, and the columns contain the number of samples classified by the discriminant analysis. Grade A is taken as an example: eight out of nine samples were classified as A and one was re-classified as C. Four samples of the 62 analyzed were re-classified in different grades, indicating that the discriminant analysis reproduced the conventional grading at 93.5%.

Table 3: Discriminant Analysis (DA) 1988 Crop. Nine Variables by Grade

Grade	Discriminant Analysis						
	A	B	C	D	E	F	G
A	8		1				
B		8		1			
C	1		8		1		
D				8			
E					8		
F						9	
G							9
Total	9	8	9	9	9	9	9
Classified by DA	8	8	8	8	8	9	9
Misclassified	1	0	1	1	1	0	0
Reproduced (%)	89	100	89	89	89	100	100

Table 4. Principal Component Analysis (PC). Component Loadings

Variable	PC1	PC2	PC3
Sugar	-0.96	0.14	-0.03
Nitrogen	0.83	0.08	0.53
Nicotine	0.14	0.98	0.05
PEE	0.60	0.18	-0.69
TVB	0.87	0.26	0.26
Ashes	0.65	-0.46	-0.51
Chlorides	0.45	-0.14	-0.26
Sug-Nic	-0.90	-0.35	0.01
Nit-Nic	0.47	-0.78	0.37
Variance (%)	48.95	22.73	14.35
Variance Accumulate (%)	48.95	71.68	86.03

2023118298

These same samples were analyzed by the Principal Component method (PC) using the correlation matrix to avoid the effect of differences in the units of measure (Table 4). The loading vectors of the components represent the correlation between each PC and the variables analyzed. PC1 includes information about nearly 50% of the system's total variance, representing mainly the variables Reducing Sugars, Nitrogen, TVB, Ashes, Chlorides and the Sug-Nic ratio; while PC2, dominated by Nicotine and the Nit-Nic ratio, represents almost 23% of the total. Thus, the nine parameters can be reduced to three new variables (PC1, PC2, PC3) involving 86% of the information contained in the original variables, significantly reducing the dimensions of the population and allowing a comprehensive analysis of the information. The discriminant analysis shown in table 3 was reproduced at 93.6% when the three PC were used (Table 5); an analysis with four PC gave the same results as with three.

Table 5. Discriminant Analysis (DA)
1988 Crop. Three Principal Components

Grade	Discriminant Analysis						
	A	B	C	D	E	F	G
A	8		2				
B		8		1			
C	1		7		1		
D				8			
E					8	1	1
F						8	
G							8
Total	9	8	9	9	9	9	9
Classified by DA	8	8	7	8	8	8	8
Misclassified	1	0	2	1	1	1	1
Reproduced (%)	89	100	78	89	89	89	89

Using PC1 and PC2, a dispersion diagram was made for the grades considered (Fig. 1); grades F and G showed the highest variability; this fact was confirmed with the results of the generalized variance. The face system developed by Riedwyl and Flury in 1986, where each element of the face represents a different variable (Fig. 2), was used with the mean vectors of each grade in their first two PC, and also with the mean vectors of the nine original variables (Fig. 3). The differences or similarities between grades (higher or lesser distance between faces) can be observed, as well as the differences between variables in each grade (differences in face's elements).

b. Comparison of Crops

Seven grades from each of the 1986, 1987 and 1988 Flue-cured crops were compared among them by means of their PC. Figure 4 shows how the grades D, E, F, G form separate groups, indicating a good correlation within a group for different crops. Grades B and C are quite near and grade A had the highest

2023118299

FIGURE 1

PRINCIPAL COMPONENT ANALYSIS

DISPERSION ANALYSIS
FLUE-CURED TOBACCO GRADES - 1988 CROP

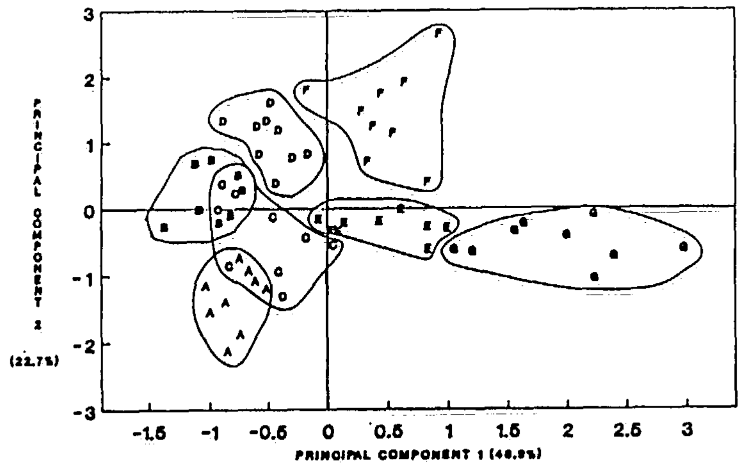


FIGURE 2. CHEMICAL VARIABLES REPRESENTED BY RIEDWYL & FLURY'S SYMETRIC FACES

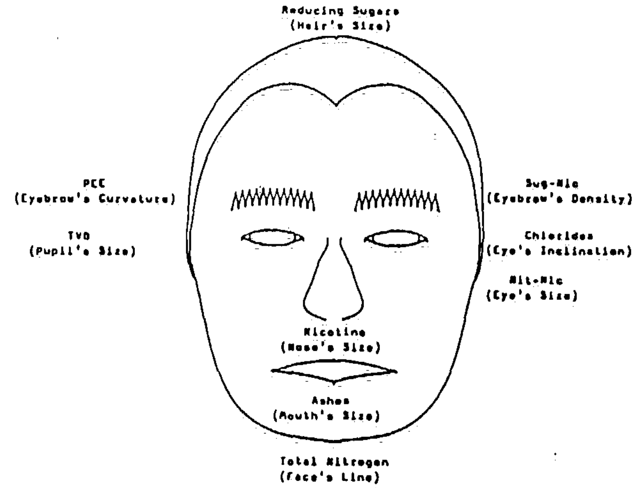


FIGURE 3

PRINCIPAL COMPONENT ANALYSIS

MEAN VECTORS
FLUE-CURED TOBACCO GRADES - 1988 CROP

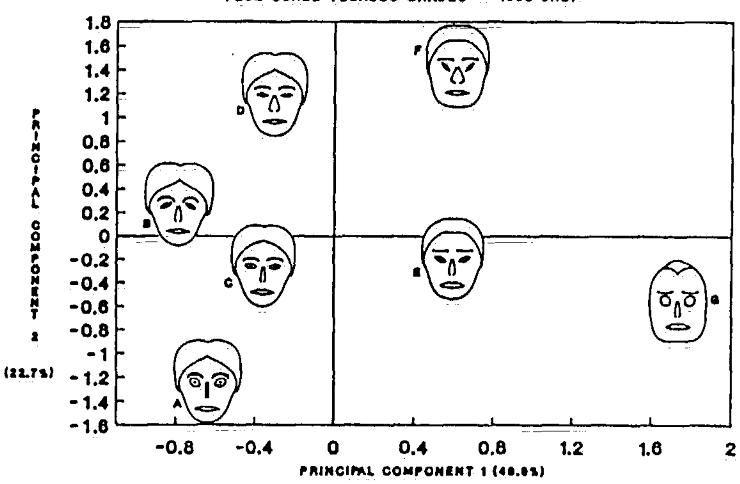
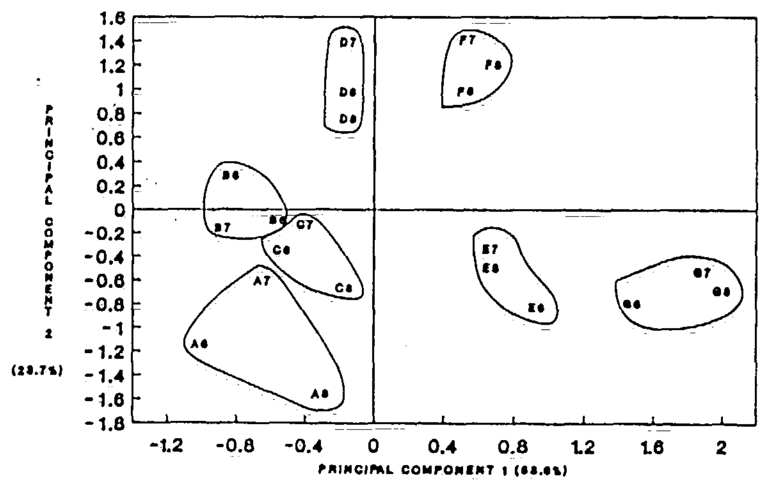


FIGURE 4

PRINCIPAL COMPONENT ANALYSIS OF FLUE-CURED TOBACCO

COMPARISON OF THE SAME GRADES BETWEEN 1986, 1987 AND 1988 CROPS



2023118300

variability in all three crops; grade A6 seems more similar to grade C than to grades A7 and A8. The mean vector of the chemical variables for each grade can be observed in Figure 5.

A generalized Hotelling test was conducted on the Mahalanobis distances among the 21 grades (Table 6), using the information from the first three PC to determine chemical equivalence between grades. Table 7 shows the results for some grades. The grade pairs D6-D8, C6-C7 and C6-A7 were equivalent among them.

The mean vectors of the three PC of each grade were analyzed with a hierarchical cluster method (Nearest Neighbor using Euclidian Distance) and the results obtained are shown on Figure 6. The groups formed by D, E, F, G confirm the consistency of these grades along the three crops. Groups A8, A6-C8-A7-C6-C7-B6, and B7-B8 show a deviation from the expected classification. A rigorous examination of these grades' samples which show inconsistency could lead to a re-classification or to a simplification or change in the grading criteria.

TABLE 6. MAHALANOBIS DISTANCE

	G8	F8	E8	D8	C8	B8	A8	G7	F7	E7	D7	C7	B7	A7	G6	F6	E6	D6	C6	B6	A6
A6	43.0	32.4	13.2	17.4	3.9	8.5	3.2	39.7	35.5	16.3	26.6	4.9	8.9	5.5	33.0	30.7	16.5	20.9	3.0	7.6	0.0
B6	42.7	13.5	14.5	3.4	5.9	1.7	13.8	38.9	14.0	12.7	6.5	1.7	2.0	3.3	32.9	12.7	21.0	4.5	2.1	0.0	
C6	31.3	15.7	7.1	6.0	1.1	2.0	5.3	28.3	17.0	6.8	12.2	0.2	3.2	0.3	23.0	14.7	11.5	8.3	0.0		
D6	34.9	3.2	16.6	0.2	11.4	3.7	24.1	31.0	2.8	11.8	0.5	5.9	5.1	11.6	26.5	2.8	22.3	0.0			
E6	7.8	19.3	0.9	18.6	5.7	15.8	7.2	5.6	24.7	1.7	28.3	11.0	20.9	11.8	4.8	19.4	0.0				
F6	22.8	4.5	16.2	3.1	15.2	8.3	28.4	20.5	0.6	9.9	3.5	11.3	11.1	10.9	17.3	0.0					
G6	1.3	18.4	8.2	23.6	15.3	22.6	10.6	1.2	23.0	5.5	32.4	21.1	28.2	24.8	0.0						
A7	33.1	20.0	7.5	8.9	1.2	3.5	3.8	30.4	22.5	8.1	16.1	1.0	4.6	0.0							
B7	39.2	13.7	16.0	4.0	6.3	0.5	12.7	36.2	13.7	12.3	8.0	2.8	0.0								
C7	29.1	12.1	6.6	4.0	1.4	1.4	6.8	26.0	14.0	5.7	9.6	0.0									
D7	40.3	3.5	21.7	1.3	16.3	6.4	31.2	34.6	2.6	16.3	0.0										
E7	9.7	10.1	1.0	9.2	3.3	8.2	7.9	7.4	13.8	0.0											
F7	28.6	0.4	20.4	3.7	19.4	10.9	34.9	25.6	0.0												
G7	0.7	20.2	9.2	28.0	19.7	29.2	24.8	0.0													
A8	25.6	30.5	6.2	19.9	2.5	10.8	0.0														
B8	32.2	10.3	13.4	2.5	4.1	0.0															
C8	22.2	16.2	3.1	0.6	0.0																
D8	31.7	3.5	13.4	0.0																	
E8	11.9	15.7	0.0																		
F8	23.3	0.0																			
G8	0.0																				

Tabla 7. Hotelling T^2 Generalyzed Multivariate Analysis of Variance

Pairs	D^2	F. Calculated
D6-D8	0.18	1.91
C6-C7	0.19	2.42
C6-A7	0.23	2.89
D6-D7	0.44	4.69
F7-F8	0.36	4.77
B7-B8	0.42	5.02
F6-F8	0.45	5.19
F6-F7	0.51	5.89
E6-E8	0.73	7.73

$F(3,143) = 3.92$

c. Classification According to Chemical Quality Parameters

A data file (Table 2) that contained the chemical analysis of 49 non-classified samples of Flue-cured, Burley, Black air-cured, and Extra air-cured tobaccos, was analyzed by a non-hierarchical cluster method, in order to evaluate the ability of this technique to group samples of similar chemical characteristics, using the first four PC containing 97.8% of the total population variance (Figure 7). Five of the six formed groups correspond clearly with the five types of tobacco, and only one sample was classified as a different group. This technique has a great potential to classify tobaccos of unknown nature, according to their chemical quality parameters.

CONCLUSION

Multivariate techniques represent a more powerful tool than the ones that have been used (univariate statistics) for the evaluation of tobacco grading systems, crops comparison and determination of equivalences among tobaccos. Although

FIGURE 5

PRINCIPAL COMPONENT ANALYSIS OF FLUE-CURED TOBACCO

GRADES OF 1986, 1987 AND 1988 CROPS

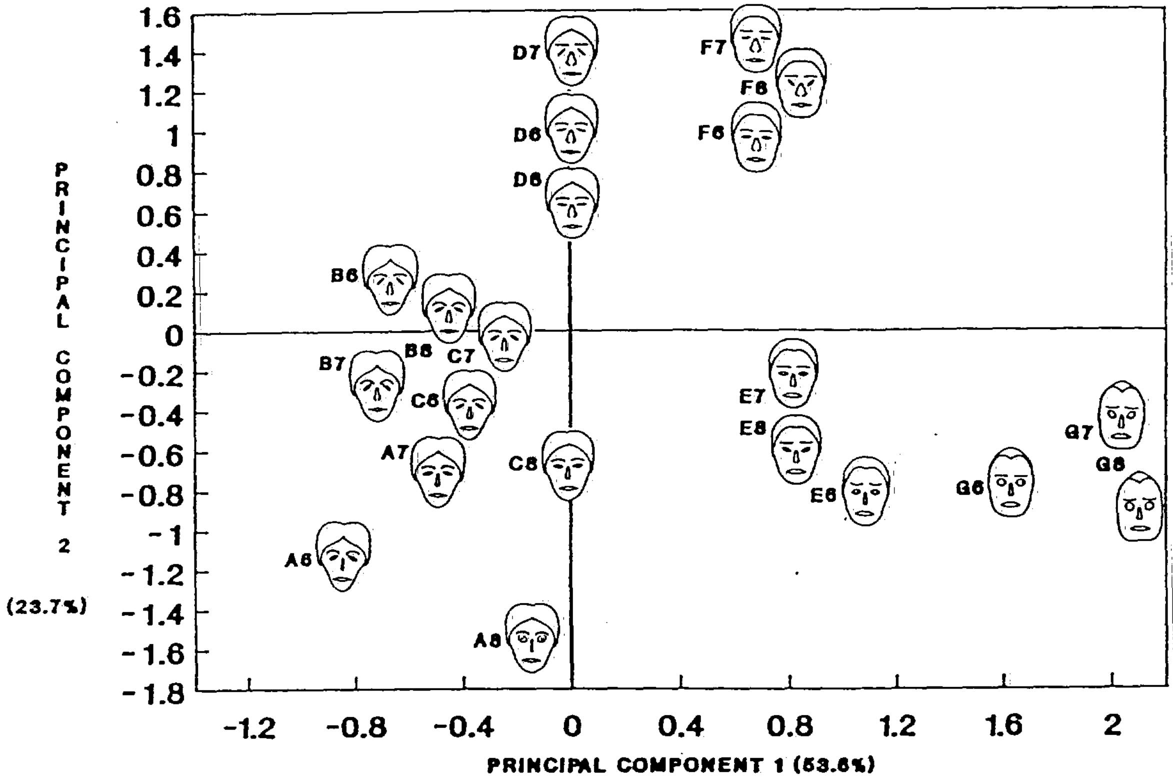
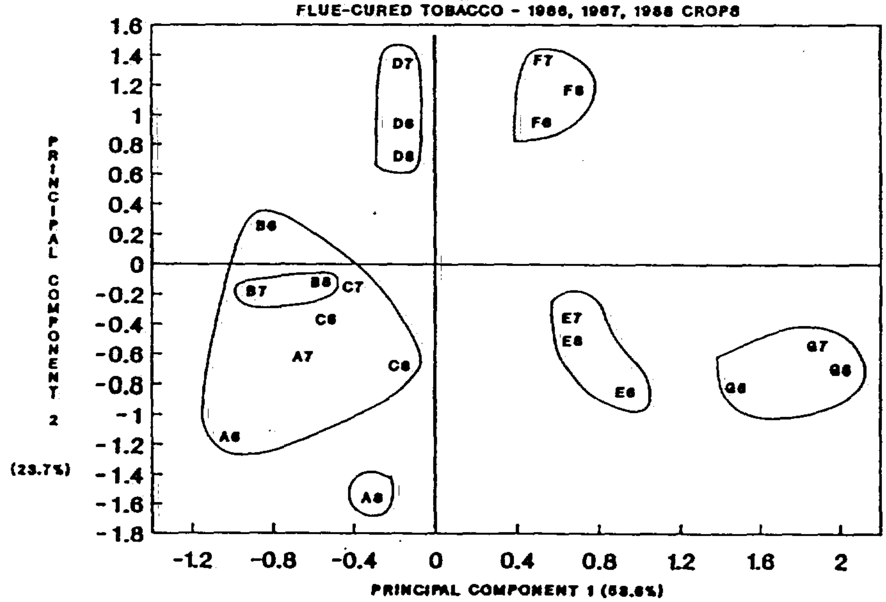


FIGURE 6

HIERARCHICAL CLUSTER ANALYSIS

(NEAREST NEIGHBOUR - EUCLIDEAN DISTANCE)

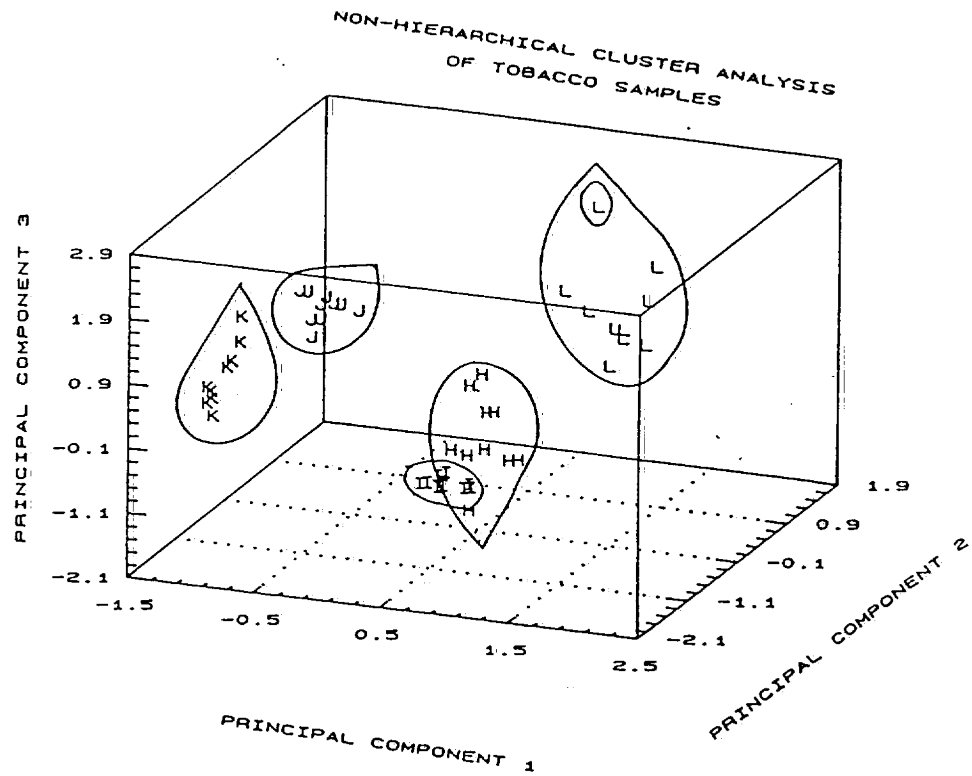
FLUE-CURED TOBACCO - 1986, 1987, 1988 CROPS



2023118302

this study only used chemical variables, the inclusion of other sensitive parameters, such as organoleptic and physical variables in addition to the chemical ones, would give a more complete information and strengthen this technique for the foregoing applications.

FIGURE 7



REFERENCES

Dimov N. and A., Tsoutsoulova. Perfu mer & Flavorist. 1987, 12 (6) 45-50.

Hsu, F.S., B.W. Good, M.E. Parrish, and T.D. Crews. Journal of HRC & CC. 1982, 5 (12) 648-655.

Jadraque, D., and B., Lopez. Spanish Burley & Bright Tobaccos: A Multivariate Statistical Approach. 9th International Tobacco Scientific Congress. 1988, Guangzhou, China.

Mardia, K.V., J.T., Kent, and J.M., Bibby. Multivariate Analysis. London, Academic Press, 1979.

Parrish, M.E. et.al. Anal. Chem. 1981, 53 (6) 826-831.

Saint-Jalm, Y. Caracterisation de Tabacs D'orient par Analyse Chromatographique D'une Fraction Volatile. Coresta Technology Group Meeting. 1989, Interlaken, Switzerland.

Saint-Jalm, Y., Testa, A. Application des Methodes de L'analyse des Données á L'analyse Chimique du Tabac et de la Fumée. 7th Coresta Sympossium. 1986, Taormina, Italy.

Schulten, H.R. Beitr. Tabakforsch. Int. 1986, 13 (5) 219-227.

2023118303